



Comparing the performance of species distribution models of *Zostera marina*: Implications for conservation[☆]



Mireia Valle^{a,*}, Marieke M. van Katwijk^b, Dick J. de Jong^c, Tjeerd J. Bouma^d, Aafke M. Schipper^b, Guillem Chust^a, Blas M. Benito^e, Joxe M. Garmendia^f, Ángel Borja^f

^a AZTI-Tecnalia, Marine Research Division, Txatxarramendi irla z/g, 48395, Sukarrieta, Spain

^b Radboud University Nijmegen, Department of Environmental Science, Institute for Water and Wetland Research, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

^c Directorate-general for Public Works and Water Management, Division Zeeland of the Ministry of the Infrastructure and Environment, P.O. Box 8039, NL-4330 EA Middelburg, The Netherlands

^d Royal Netherlands Institute for Sea Research (NIOZ), P.O. Box 140, 4400AC Yerseke, The Netherlands

^e University of Granada, Department of Ecology, Andalusian Environmental Centre, Avda. del Mediterráneo s/n, 18006, Granada, Spain

^f AZTI-Tecnalia, Marine Research Division, Herrera Kaia, Portualdea z/g, E-20110 Pasaia, Spain

ARTICLE INFO

Article history:

Received 13 December 2012

Received in revised form 5 March 2013

Accepted 8 March 2013

Available online 15 March 2013

Keywords:

Intertidal

Dynamics

Conservation

Seagrasses

Ecosystem Management

Wadden Sea

ABSTRACT

Intertidal seagrasses show high variability in their extent and location, with local extinctions and (re-)colonizations being inherent in their population dynamics. Suitable habitats are identified usually using Species Distribution Models (SDM), based upon the overall distribution of the species; thus, accounting solely for spatial variability. To include temporal effects caused by large interannual variability, we constructed SDMs for different combinations and fusions of yearly distribution data. The main objectives were to: (i) assess the spatio-temporal dynamics of an intertidal seagrass bed of *Zostera marina*; (ii) select the most accurate SDM techniques to model different temporal distribution data subsets of the species; (iii) assess the relative importance of the environmental variables for each data subset; and (iv) evaluate the accuracy of the models to predict species conservation areas, addressing implications for management. To address these objectives, a time series of 14-year distribution data of *Zostera marina* in the Ems estuary (The Netherlands) was used to build different data subsets: (1) total presence area; (2) a conservative estimate of the total presence area, defined as the area which had been occupied during at least 4 years; (3) core area, defined as the area which had been occupied during at least 2/3 of the total period; and (4–6) three random selections of monitoring years. On average, colonized and disappeared areas of the species in the Ems estuary showed remarkably similar transition probabilities of 12.7% and 12.9%, respectively. SDMs based upon machine-learning methods (Boosted Regression Trees and Random Forest) outperformed regression-based methods. Current velocity and wave exposure were the most important variables predicting the species presence for widely distributed data. Depth and sea floor slope were relevant to predict conservative presence area and core area. It is concluded that, the fusion of the spatial distribution data from four monitoring years could be enough to establish an accurate habitat suitability model of *Zostera marina* in the Ems estuary. The methodology presented offers a promising tool for selecting realistic conservation areas for those species that show high population dynamics, such as many estuarine and coastal species.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Seagrasses play an important role in maintaining a diverse and healthy coastal ecosystem (Björk et al., 2008) and providing many environmental functions, which lead seagrass ecosystems to be amongst the

most valuable ecosystems in the world (Costanza et al., 1997). However, their habitat is being fragmented and lost worldwide (Duarte, 2002; Hughes et al., 2009), with rates of decline accelerating in recent years (Waycott et al., 2009), and seagrass beds disappearing completely in some areas (Green and Short, 2003; Kirkman, 1997; Short et al., 2006). In contrast to this global crisis of seagrass ecosystems, recent researches have detected a recovery of mixed intertidal beds of *Zostera marina* and *Zostera noltii* in the North Frisian Wadden Sea (Germany) (Dolch et al., 2012), and a steady and linear increase in *Z. noltii* meadow areas within Bourgneuf Bay (France) (Barillé et al., 2010). These encouraging results reveal the potential for seagrass recovery, highlighting the importance of the assignment of suitable areas to permit the conservation of these valuable ecosystems.

Abbreviations: SDM, Species Distribution Models; TPA, total presence area; CPA, conservative presence area.

[☆] Given his role as Managing Guest Editor, Ángel Borja had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to Joao-Carlos Marques.

* Corresponding author. Tel.: +34 667174451.

E-mail address: mvalle@azti.es (M. Valle).

Seagrass meadows are highly dynamic ecosystems, maintained through the continuous recruitment of new clones, in combination with the growth and turnover of the shoots (den Hartog, 1971; Duarte et al., 2006). In semi-annual populations, seed production, dispersal, germination and seedling survival additionally determine the bed dynamics. In particular, the intertidal habitat-forming species, such as the semi-annual flexible type of *Z. marina*, but also the perennial *Z. noltii*, have high inter annual variability in extent and location, with local extinction and (re-)colonizations being part of their life strategy (Erfteimeijer, 2005; van Katwijk et al., 2006, 2009). Consequently, not all suitable habitats of the species are occupied every year. However, these uninhabited, but suitable, areas are important for seagrasses to survive in the long-term, by providing refuge areas to overcome temporarily unsuitable local circumstances elsewhere, for example, to overcome weather conditions such as the presence of ice. Loss of these unoccupied habitats would decrease the possibilities for seagrass survival, enhancing ultimately the risk of extinction. Even if *Zostera* species and their habitats are protected under one or more environmental frameworks at international, European and national scales (Bos et al., 2005; Valle et al., 2011), the protection measures are based often upon the actual distribution of the species in a particular monitored year, rather than the overall available habitat. In this way, human activities (such as: bottom trawling; shellfish and worm collecting; and recreational activities) remain permitted around the seagrass beds. Such activities may lead to damage of the unoccupied habitat of these dynamic populations (Cunha et al., 2012).

Areas that are occupied frequently by seagrass (core areas, hereafter) can be distinguished easily when a time-series of monitoring data is available (Dolch et al., 2012). When seagrass monitoring is infrequent, it is likely to incorporate the core areas, but likely also to exclude large areas that are occupied only occasionally; thereby, underestimating the overall seagrass habitat. If only core areas are protected, a large part of the total distribution could be lost, e.g. in The Netherlands, this amounts to $91\% \pm 8.59\%$ (Table 1). Therefore, defining the entire potential habitat area needed to maximally protect a species with high temporal dynamics requires frequent monitoring of its distribution. As the seagrass monitoring is laborious and expensive, there is a need for other approaches to be adopted for the delimitation of the conservation areas. One approach could be the use of species distribution models (sensu Guisan and Zimmermann, 2000; SDM, hereafter), to identify suitable seagrass habitats.

In recent years, several techniques to build SDMs have been developed (Elith and Leathwick, 2009; Franklin, 2009; Guisan and Thuiller, 2005). SDMs are based commonly on the overall distribution of a species, without considering if there are ecological differences between frequently and occasionally occupied areas. Likewise, distribution patterns

of highly dynamic species (such as seagrasses) may vary considerably between different years of monitoring. In some years, seagrasses may spread to areas which are normally not occupied. Incorporation of such incidental occurrences in a SDM is likely to decrease the robustness of the model. On the other hand, seagrass beds could be damaged or represent only core areas, because of insufficient protection. Models based upon these frequently occupied areas might represent only a part of the total area suitable for the species. Therefore, in the research developed here, different subsets of a time-series of monitoring data were used to build SDMs for a seagrass species. Using a 14-year time-series of distribution data of *Z. marina* in the Ems estuary (The Netherlands), which allows addressing spatio-temporal variability in the modelling, the aims of this research are to: (i) assess the spatio-temporal dynamics of an intertidal seagrass bed of *Zostera marina*; (ii) select the most accurate SDM techniques to model different temporal distribution data subsets of the species; (iii) assess the relative importance of the environmental variables for each data subset; and (iv) evaluate the accuracy of the models to predict species conservation areas, addressing implications for management.

2. Material and methods

2.1. Study area

The Wadden Sea (Fig. 1a) is one of the world's largest international marine wetland reserves (approx. 6000 km²), bordering the coasts of The Netherlands, Germany and Denmark. Due to its high ecological importance, it is under the protection and conservation frameworks of three main European Directives: the Habitat Directive (92/43/EEC), the Directive on the Conservation of Wild Birds (2009/147/EC); and the Water Framework Directive (2000/60/EC).

The Ems estuary (Fig. 1b), located on the border between The Netherlands and Germany, is one of the most important estuaries intersecting the Dutch Wadden Sea (de Jonge, 2000). The biological and physical processes affecting this estuary have been investigated extensively (Baretta and Ruardij, 1988; de Jonge, 1992a, 1992b, 2000). Dominant physical processes are the tides (range and currents), wind-generated waves, and freshwater inflow from the Ems river and the Westerwoldse Aa river (Fig. 1) (Talke and de Swart, 2006). The availability of a time-series of 14 years on species' distribution and the relatively stable presence of *Z. marina* populations in this area (Bos et al., 2005) during the studied period, together with environmental data availability, were the main reasons to select the intertidal area of this estuary as the study area (Fig. 1c) to undertake this research.

Table 1

Occupied area, core area and potential loss of 14 seagrass beds in The Netherlands. Occupied area is the total area occupied during any of the monitoring years. Core areas are areas with an occupation of equal or greater than 66% of the monitored years. Potential seagrass bed loss is the potential loss if only core areas are protected (data courtesy: Ministry of Transport, Water Management and Public Works (analyses: Annette Wielemaker)).

Water body	Species	Location	Period	Monitoring years	Occupied area (ha)	Core areas (ha)	Potential loss (%)
Wadden Sea	<i>Z. noltii</i>	Groningse kwelders	1991–2008	13	115.94	14.05	88
		Tersch-Oosterend	1991–2008	14	50.22	8.25	84
	<i>Z. marina</i>	Ems estuary	1991–2008	14	251.24	15.7	94
		Tersch-haven	1991–2001	7	25.13	1.71	93
Eastern Scheldt	<i>Z. noltii</i>	Dortsman	1990–2009	10	58.08	15.09	74
		Kats	1991–2008	10	13.94	2.57	82
		Kattendijke	1990–2009	8	13.89	2.08	85
		Krabbekreek Noord	1990–2009	12	36.42	0.29	99
		Sint Annaland	1990–2009	12	3.35	0	100
		Mastgat	1990–2009	11	4.6	0	100
		Viane	1992–2008	8	19.37	1.22	94
		Zandkreek Noord	1990–2009	18	34.55	0.1	100
		Zandkreek Zuid	1990–2009	18	37.52	7.63	80
		Zuid Beveland	1990–2009	11	34.37	1.18	97

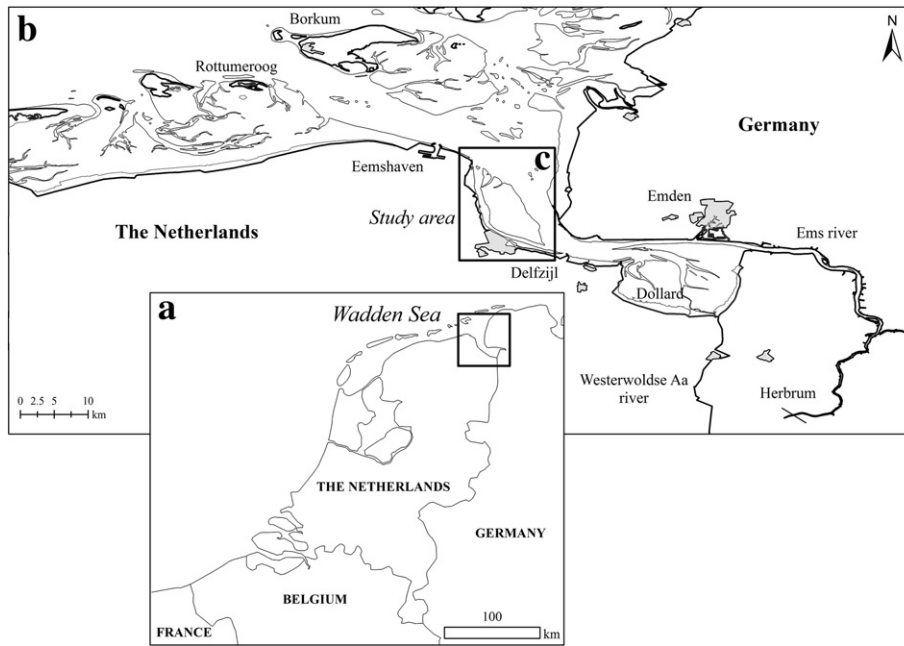


Fig. 1. (a) Map of the Wadden Sea (North-West Europe) with inset (b) of the Ems estuary and (c) the study area.

2.2. Data collection and processing

2.2.1. Species presence data

The seagrass beds in the Wadden Sea have been monitored on a yearly basis since the mid-1990s, by the Directorate-general for Public Works, under the framework of the biological monitoring programme

of the Dutch Government. Monitoring has been carried out based upon both aerial photographs and ground surveys. In the ground surveys the entire potential seagrass area was visited; thus, the mapping was as complete as possible. Monitored data of *Z. marina* in the Ems estuary from 1995 to 2009 (with the exception of 1998, which was not monitored) were used to determine the species' frequency of occurrence by

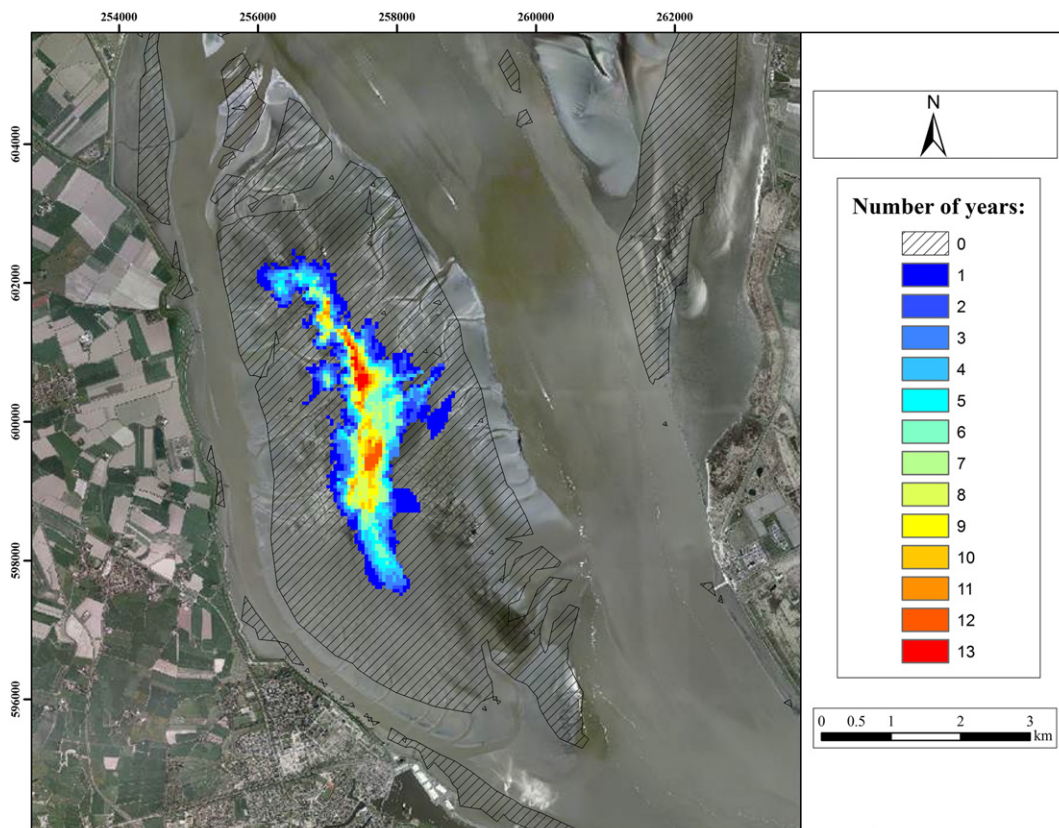


Fig. 2. Distribution map of *Zostera marina* in the Ems estuary showing the frequency of occurrence (number of years) within the period 1995–2009 (excluding 1998).

overlaying the yearly spatial distribution maps (Fig. 2). Based on this frequency map, six different presence data subsets were built: (1) total presence area (TPA), as represented by an occupation frequency greater than 0; (2) a conservative estimate of the total presence area (hereafter: Conservative Presence Area, CPA), defined as the area which had been occupied over at least 4 years, thus excluding occasionally occupied areas; (3) core area, defined as the area which had been occupied during at least 2/3 of the total period (i.e. at least 9 years); and (4–6) three random selections of four monitoring years. Using four monitoring years enabled an assessment of whether low-frequency monitoring is sufficient to establish a representative SDM, as seagrass monitoring is expensive and less intensive monitoring could reduce the costs.

2.2.2. Environmental data

Environmental variables that are known to affect the species' distribution (Bos et al., 2005) were provided by the Ministry of Transport, Water Management and Public Works. The variables included current velocity, wave exposure, depth, sea floor slope, and salinity in the water column (Table 2). Current velocity and wave exposure maps were created using a coastal zone model for currents (Simulating WAVes Nearshore, SWAN), assuming dominant wind directions and moderate storm velocities (Bos et al., 2005). Depth data were obtained from the bathymetric map of the Wadden Sea (2002, Ministry of Transport, Water Management and Public Works). Sea floor slope was derived from depth, using the Spatial Analyst 9.3 extension from ArcGIS 9.2 software (ESRI®). Salinity was derived from a 2D model, based on water movements and quality, and local freshwater sources (Jager and Bartels, 2002). All of the variables were masked to the same extent and exported as raster-based grid files with the same spatial resolution (50 m).

2.3. Analysis of spatio-temporal dynamics

Spatio-temporal dynamics (cycles of decline and recovery) of the species were analyzed by computing differences in vegetated area between couples of subsequent monitoring years, using ArcGIS 9.2 software (ESRI®). Four change classes were defined: (i) colonized (area newly vegetated, in comparison with the preceding monitoring year); (ii) stable (area vegetated at present, as well as during the preceding monitoring year); (iii) disappeared (area vegetated in the preceding monitoring year, but not in the present year); and (iv) non-occupied (area unvegetated during both of the monitoring years). The probability of change from state *i* (present or absent), to state *j* (present or absent), in a time-step *t* was estimated as:

$$P_{i(t) \rightarrow j(t+1)} = \frac{N_{i(t)j(t+1)}}{\sum N} \tag{1}$$

where $N_{i(t)j(t+1)}$ is the number of pixels at state *i* in year *t* that shift to state *j* at year *t* + 1; and $\sum N$ is the total number of pixels at all states. Based upon these probability values, a transition probability matrix was established.

Table 2
Ranges in environmental variables values: Min—minimum; Max—maximum; Mean—mean; and SD—standard deviation.

Type	Variable	Min	Max	Mean	SD	Units
Water dynamics	Current velocity	0.16	1.65	0.55	0.15	m s ⁻¹
	Wave exposure	0.09	0.54	0.33	0.08	m s ⁻¹
Topography	Depth	-647	279	-57.96	64.28	cm
	Sea floor slope	0.00	79.97	12.03	12.63	degrees
Water characteristics	Salinity	2.30	28.80	21.12	2.58	PSU

2.4. Habitat modelling

In order to obtain accurate predictions, an adequate selection of the modelling algorithm is critical. As an arbitrary selection of a single modelling algorithm could lead to sub-optimal results (Elith et al., 2006), multiple SDMs were built for each of the six presence data subsets using the predictor variables included in Table 2. Eight different modelling techniques were applied, in order to determine the best modelling technique for each presence data subset. The techniques adopted included 5 machine-learning methods and 3 regression-based methods (Table 3). Machine-learning methods include different algorithms that derive the mapping function or classification rules inductively, directly from the training data (Breiman, 2010; Gahegan, 2003). Regression-based models include a Generalized Linear Model (GLM) and its 'non-parametric' extensions (Guisan et al., 2002).

For modelling purposes, pseudo-absences were created along the never occupied area, avoiding spatial overlap with the presences. Entering the different presence data subsets as a dependent variable and the selected environmental variables as the predictors, spatial distribution modelling was performed.

Area Under the receiver operating characteristic Curve (AUC) was selected as an evaluation method (Fielding and Bell, 1997), since it is amongst the most widely-applied measures of model accuracy within the context of SDMs (Raes and ter Steege, 2007). The AUC provides a summary measure of model discrimination accuracy (Pearce and Ferrier, 2000) being equivalent to the probability that a model ranks a randomly-selected presence site, higher than a randomly-selected absence site (Liu et al., 2011). AUC values range between 0.5 (random sorting) and 1 (perfect discrimination). This method is known to be an appropriate and useful way to summarize model performance, especially when comparing different modelling methods (Franklin, 2009; Lobo et al., 2008). The AUC values for each model were computed using k-fold validation (5 groups), with the 'evaluate' function of the R package 'dismo' (Hijmans et al., 2012).

Variable importance analysis was computed for each selected model applying the specific functions from their corresponding packages (Table 3). Relative importance of variables is obtained as a percentage for machine-learning based methods. Thus, in order to compare resulting rankings, with regression-based methods results, the obtained standardized regression coefficients were rescaled to percentages.

2.5. External evaluation of the models

In order to perform the external evaluation of the models, CPA was considered as the reference area to be predicted; this area excludes the occasional presences and could be considered as the most accurate approximation of the entire potentially-suitable habitat. Firstly,

Table 3
Overview of techniques applied to establish species distribution models of the different presence data subsets of *Zostera marina* in the Ems estuary.

Group of SDMs	Id	Technique	Platform	References
Machine learning methods	1	Boosted Regression Trees	R package: dismo	Hijmans et al., 2012
	2	MaxEnt	MaxEnt software	Phillips et al., 2006
	3	Artificial Neural Networks	R package: nnet	Venables and Ripley, 2002
	4	Random Forest	R package: randomForest	Liaw and Wiener, 2002
Regression-based models	5	Support Vector Machines	R package: kernlab	Karatzoglou et al., 2004
	6	Generalized Additive Models	R package: gam	Hastie, 2011
	7	Generalized Linear Models	R package: dismo	Hijmans et al., 2012
	8	Multivariate Adaptive Regression Splines	R package: earth	Milborrow et al., 2012

in order to convert the continuous probability maps generated by the models (with values ranging from 0 to 100) into binary predictions maps (suitable and unsuitable habitat), probability thresholds were identified. Threshold values were obtained with the 'evaluate' function of the R package 'dismo' (Hijmans et al., 2012), maximizing sensitivity (true positive rate: TPR, hereafter) and specificity (true negative rate: TNR, hereafter). In the case of the core area model, two thresholds were identified to establish the binary maps: accurate (core model 1: maximizing TPR and TNR) and broad (core model 2: threshold fitted in 25% of the probability). Under this second threshold, areas with a predicted probability greater than 25% were considered as being suitable areas. For evaluation, binary prediction maps were compared to the CPA. The TPA model was not used in the model evaluation, since it was built based upon a larger presence area than the CPA. Four different classes were distinguished in the resulting maps: (i) never vegetated and predicted as unsuitable (X_{11}); (ii) never vegetated and predicted as suitable (X_{12}); (iii) vegetated and predicted as unsuitable (X_{21}); and (iv) vegetated and predicted as suitable (X_{22}). A confusion matrix (also called an error matrix) was built for each model based on the area (in hectares) occupied by each of the above defined classes (Table 4). Different accuracy measures can be computed from the confusion matrix (Congalton, 1991). In order to assess the prediction accuracy of the selected models, omission and commission errors were calculated. The error values range between 0 (no error) and 1 (highest error). Omission error, defined as the exclusion error or underestimation, is indicative of the accuracy of the model classifying observed categories. High omission error (values greater than 0.5) implies low accuracy, since the model is not able to predict the observed presences as being suitable areas. Commission error, defined as the inclusion error or overestimation, could be interpreted as being a potential area that is not occupied by the species due to the dispersal limitation of the species, or other environmental factors which are limiting the dispersion (Guisan and Zimmermann, 2000). Therefore, the relative cost of underestimation is higher than that of overestimation (Fielding and Bell, 1997).

3. Results

3.1. Dynamics of the intertidal seagrass beds

Interannual changes in seagrass presence were divided into: colonized (i.e. area newly vegetated, in comparison with the preceding monitoring year); stable (i.e. area vegetated at present, as well as during the preceding monitoring year); and disappeared (i.e. area vegetated in the preceding monitoring year, but not in the present year) (Fig. 3). From 2001 to 2007, the population maintained a stable surface area of around 150 ha. On average, colonized and disappeared areas showed remarkably similar transition probabilities of 12.7% and 12.9%, respectively (Table 5). Stable presence areas made up 20% of the total presence area, whereas stable unvegetated areas accounted for 54.3% (Table 5).

Table 4
Components of the confusion matrix.

		Model			
		Unsuitable	Suitable	Subtotal	Omission error
Observed	Never vegetated	X_{11}	X_{12}	$\sum X_{1i}$	$1 - (X_{11}/\sum X_{1i})$
	Vegetated	X_{21}	X_{22}	$\sum X_{2i}$	$1 - (X_{22}/\sum X_{2i})$
	Subtotal	$\sum X_{i1}$	$\sum X_{i2}$		
	Commission error	$1 - (X_{11}/\sum X_{i1})$	$1 - (X_{22}/\sum X_{i2})$		

3.2. Best SDM technique selection

Based upon the AUC evaluation method, a consistently high predictive accuracy was found for the SDMs built on the different presence data subsets (Table S1). The machine-learning methods: Boosted Regression Trees (BRT) and Random Forest (RF), were the best performing techniques; in contrast, GLM showed the lowest AUC values (Table S1). BRT provided the most accurate models for TPA and CPA; and RF for core area and the three random selections of four monitoring years.

3.3. Variable relative importance on selected SDMs

For all presence data subsets, current velocity was the most important variable determining the species distribution. This variable was followed by wave exposure and depth for the TPA and the three random subsets models (Table 6). The ranking of the predictor variables' importance was different for the CPA and the core area subsets models. The CPA model ranked depth as a more important variable than wave exposure. For the core area model, the importance of sea floor slope increased considerably, it was the second most important variable explaining the species distribution, followed by depth (Table 6).

3.4. Model external evaluation

Predicted suitable areas presented a similar spatial pattern (Fig. 4). All of the maps predicted most of the observed presence areas as being suitable, but also predicted some areas to be suitable where the species was not observed, such as the areas to the east and west of the CPA. Suitable areas were more limited in the case of the core area models and they were more extended in all other cases.

Judged from the omission and commission errors' calculations (Table S2), core area models were the least accurate models predicting the CPA: they showed the highest omission errors, but presented the lowest commission errors. When the broad threshold was applied to the core area model (threshold fitted in 25% of the probability), the resulting binary map (core 2) showed a lower omission error. All other models had a very low omission error, whereas the commission error was relatively high.

4. Discussion

4.1. Spatio-temporal dynamics of an intertidal seagrass bed

During the 14 years studied, *Z. marina* beds in the Ems estuary showed a remarkably similar transition probability of colonized and disappeared areas: on average, 12.9% of the vegetated areas turned

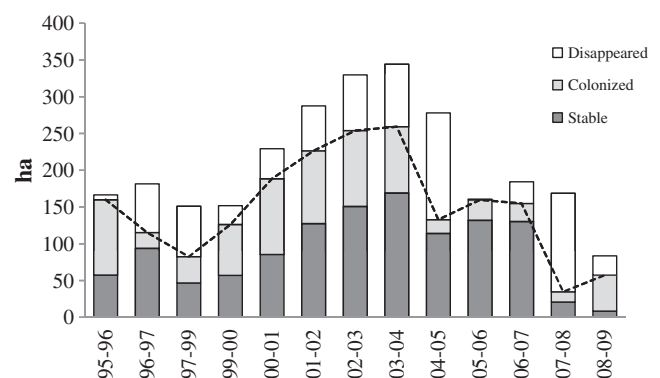


Fig. 3. Dynamics of *Zostera marina* beds in the Ems estuary: colonized (area newly vegetated, in comparison with the preceding monitoring year); stable (area vegetated at present, as well as during the preceding monitoring year); and disappeared (area vegetated in the preceding monitoring year, but not in the present year). Dashed line shows the total area covered by the species in the second year mentioned in each bar.

Table 5

Transition probability matrix showing surface mean and standard deviation values in ha and in percentages per each change class (non-occupied, colonized, disappeared and stable).

		Year + 1	
		Absence	Presence
Year	Absence	248.77 ± 78.38 54.30% (Non-occupied)	58.35 ± 36.87 12.70% (Colonized)
	Presence	58.87 ± 44.43 12.90% (Disappeared)	91.77 ± 50.60 20.00% (Stable)

into a bare state each year whereas, on average, 12.7% of the bare areas became colonized each year. This pattern confirms the dynamic population strategy of the species: this is probably the best strategy in a dynamic environment, such as the Wadden Sea. This high reproduction or regrowth rate, which is characteristic of *r*-strategy species (MacArthur and Wilson, 1967), enables the species to recover from disturbance and is encountered in frequently-disturbed sites (Trémolières, 2004). The dynamics of the Ems estuary population are consistent with the relatively high yearly fluctuations of eelgrass presence in other Dutch and German intertidal seagrass populations (Bos et al., 2005; Dolch et al., 2012) (Table 1). Also, in other parts of the world, seagrass beds are reported to be dynamic: contours of beds may shift; extents may 'pulse'; likewise, new beds may form where others disappear (e.g., den Hartog, 1971; Krause-Jensen et al., 2004; Martinet, 1782). This observation emphasizes the importance of protecting the suitable eelgrass habitats in addition to existing eelgrass beds, for which we evaluated modelling techniques in this research.

4.2. Comparison between SDMs techniques

Many authors have performed comparisons between multiple SDM techniques (Elith et al., 2006; Opper et al., 2012; Segurado and Araújo, 2004; Zaniewski et al., 2002). Results from the present study have shown a better performance of the machine-learning methods in comparison with regression-based methods. The GLM technique performed particularly poorly for all presence data subsets. For the TPA and the CPA models, BRT obtained the highest AUC values; whilst RF produced the highest AUC for the core area model and the three models based upon random selections of four monitoring years (Table S1). BRT technique combines two algorithms: the boosting algorithm which iteratively calls the regression-tree algorithm to construct a combination or 'ensemble' of trees (Elith et al., 2006). The RF algorithm is able to fit complex non-linear surfaces from high-dimensional input data (Cutler et al., 2007). The use of both the BRT and RF techniques is presently becoming increasingly popular (Bisrat et al., 2012; Elith et al., 2006;

Table 6

Variable relative importance ranking for each of the selected models. The relative importance value of the variable is shown within parentheses (in %). Key: TPA—Total Presence Area; CPA—Conservative Presence Area; CORE—Core area; R1—Random selection 1; R2—Random selection 2; R3—Random selection 3.

TPA	CPA	CORE
Current velocity (40.03)	Current velocity (33.99)	Current velocity (22.65)
Wave exposure (30.49)	Depth (31.75)	Sea floor slope (22.43)
Depth (19.02)	Wave exposure (25.92)	Depth (21.04)
Salinity (8.37)	Sea floor slope (5.11)	Wave exposure (17.94)
Sea floor slope (2.09)	Salinity (3.23)	Salinity (15.95)
R1	R2	R3
Current velocity (28.60)	Current velocity (26.57)	Current velocity (28.06)
Wave exposure (21.89)	Wave exposure (22.30)	Wave exposure (22.34)
Depth (19.64)	Depth (21.96)	Depth (22.12)
Sea floor slope (16.92)	Salinity (15.39)	Salinity (14.23)
Salinity (12.94)	Sea floor slope (13.77)	Sea floor slope (13.25)

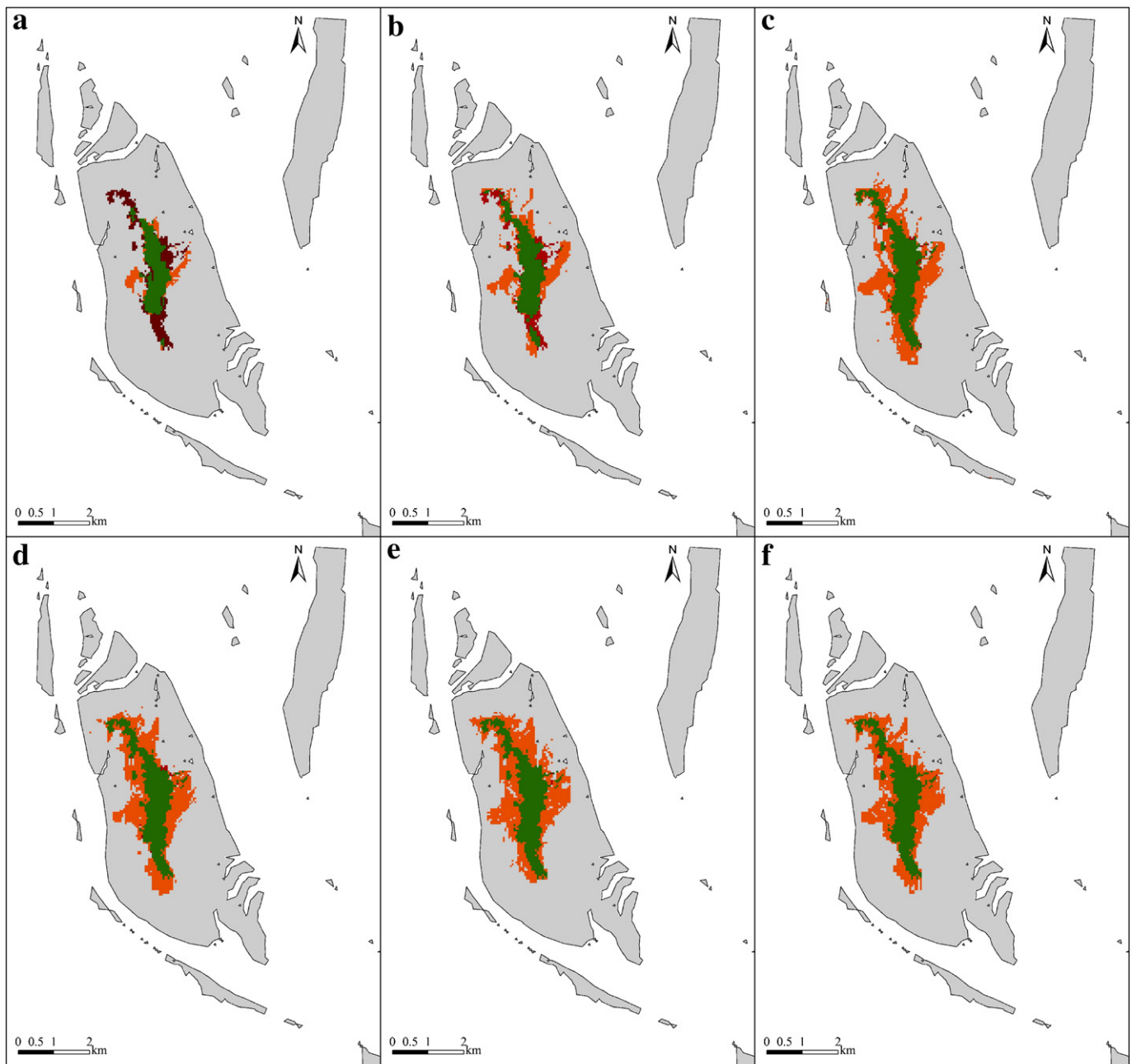
Williams et al., 2009). These machine-learning methods provide a number of advantages over the more traditional GLM approach, including: robust parameter estimates; model structure learned from data; and easy implementation of complex interactions.

4.3. Differences in variables' relative importance

Current velocity and wave exposure were found to be the most important variables predicting the species presence for TPA and for the three randomly-selected years data subsets models, whereas depth and sea floor slope were identified as being important variables to predict CPA and core area. Depth was more relevant in predicting the CPA and core area, than predicting TPA and the three random selections of four monitoring years. This outcome could be related to the more specific habitat requirements of the species (narrower depth range) in stable areas, such as CPA and core area; in contrast to the broader depth range associated with the incidental presences included in the TPA and in the three random selections of four monitoring years. Sea floor slope acquired higher importance for predicting the core area, probably because the core area is restricted largely to flat intertidal areas. The sea floor slope is the consequence of the geomorphology of the intertidal flat and coincides with higher local wave impact; this is partly a consequence of the slope, but partly due to the vicinity of the deep channels which permits the buildup of the waves. The presence points included in the models built for the three data subsets of four randomly-selected years had a wider distribution; this probably explains the documented similarity with the variables' importance in the TPA model. The high importance of current velocity for all scenarios, is consistent with the well-documented influence of the water dynamics on seagrass distribution (e.g. Bos and van Katwijk, 2007; Fonseca and Bell, 1998; van Katwijk and Hermus, 2000). Wave action has severe effects on seagrasses due to the continuous drag force on the leaves (Bos et al., 2005), together with resuspension of the sediments and subsequent burial of seagrasses (Bell et al., 1999; Fonseca et al., 2000; Han et al., 2012). Increased erosion and sedimentation was found also to affect negatively *Z. marina* settlement in the Dutch Wadden Sea (van Katwijk et al., 2000). Depth is important, as it determines the availability of light, which is essential to drive photosynthesis and, hence, growth (Greve and Binzer, 2004); however, it is also related directly to duration of exposure to rehydration, which defines the upper and lower limits of seagrass zonation (Bos et al., 2005). The semi-annual flexible type of *Z. marina* grows in the mid-intertidal zone, being less susceptible to desiccation than the perennial morphotype with stiffer sheaths, which grows in the subtidal zone (Keddy and Patriquin, 1978). In the Wadden Sea, the lower depth limit of the species studied was found to be defined directly by water dynamics (van Katwijk and Hermus, 2000). In the present work, this is supported by the higher importance of depth in predicting the CPA and the core area. Sea floor slope has been identified previously by other authors as an important regulating factor for seagrass distribution (Bekkby et al., 2008; Narumalani et al., 1997). The negative influence of high salinity has been acknowledged by many authors (Bos et al., 2005; Kamermans et al., 1999; van Katwijk et al., 1999), decreasing productivity and vitality.

4.4. Implications for conservation: monitoring frequency and effort

In general, for the designation of protected areas, a low omission error is critical to allow for the maximal recovery potential of seagrass. On the other hand, low commission error is also recommendable, in order to not unnecessarily frustrate other users (i.e. shell collectors, bottom trawling fishermen, tourists) in the area. It is a challenge to coastal managers to balance between omission and commission errors when assigning the area that needs to be protected. We applied models using core areas and areas occupied in four randomly-selected monitoring years, to predict a conservative estimate of the presence area, i.e. the area occupied by seagrass during, at least, 4 years. The core area models



Legend: Never vegetated- predicted Unsuitable Vegetated- predicted Unsuitable Never vegetated- predicted Suitable Vegetated- predicted Suitable

Fig. 4. Evaluation maps for: (a) core area model 1, threshold TPR/TNR; (b) core area model 2, threshold 25%; (c) conservative presence area model, threshold TPR/TNR; (d) random year selection 1, threshold TPR/TNR; (e) random year selection 2, threshold TPR/TNR; and (f) random year selection 3, threshold TPR/TNR.

predicted a large area as being unsuitable which was observed to be occupied frequently by the species (high omission error). The core area model was based on a limited area; thus, the resulting area predicted as being suitable was also limited. The models based on four random years' selections had a much lower omission error, with only a slightly higher commission error. Surprisingly, they performed equally well as, or even better than, the model based on the CPA.

Guisan and Zimmermann (2000) summarized that omission and commission errors in species distribution models, could result: from the low relevance of the environmental variables used in the model; from algorithmic errors; or from biological errors. This observation must be considered when interpreting the results obtained. Environmental variables must be representative of the period studied and a proper selection is critical (Elith et al., 2006). The present results

indicate that a proper prediction of the habitat area that needs to be protected requires only a limited number of monitoring years; likewise, that these years do not have to be consecutive; this suggests that intensive monitoring is not necessary. For the studied case of the Ems estuary, spatial distribution data from four monitoring years could be enough to establish an accurate habitat suitability model of *Z. marina*. Models based upon four monitoring years' data subsets rendered an equal or even better suitability model, as compared to the models based upon all of the monitoring data (14 monitoring years in total). Applying the model to a similar area, e.g. the Wadden Sea east, could assist in determining the extent to which this minimum necessary monitoring frequency is comparable between sites. It should be noted that, monitoring may serve more purposes than only establishing the area that needs to be protected.

Indeed, annual mapping is highly valuable for deriving a mechanistic understanding of the “wax and wane” of the seagrass, especially in a situation with large changes and under a gradual decline. Omission and commission errors obtained in the case of the core area model showed that the reduction of seagrass beds to core areas will reduce severely the habitat suitability prediction capacity. Management measures of those affected beds may include closure of a larger area surrounding the bed.

In conclusion, the prediction of suitable habitats for *Z. marina* species in the Ems estuary can be computed based upon relatively few environmental variables. Machine-learning methods outperformed regression-based methods, showing a higher prediction power. When limited areas (core areas) were modelled, the prediction accuracy of the potential suitable habitats decreased. Prediction of suitable habitats improved with SDMs based upon distribution data from, at least, four years of monitoring. The methodology presented offers a promising tool for selecting realistic conservation areas for those species that show high population dynamics, such as many estuarine and coastal species. The suitable habitats for the species, thus for protection, can be established by applying SDMs based upon a few years of monitoring. This approach is particularly important for target or key species which are protected under environmental frameworks (e.g. Habitat Directive (92/43/EEC)), but could also be applied to economically-valuable species. The method would benefit from further testing on other species, with available data for similar long-term monitoring, to determine: (i) how the minimal monitoring period depends upon the dynamics of the species; and (ii) to which extent the output of the method depends upon the clustering of the species. Considering the difficulties to overcome in the protection of habitats that could be temporally unoccupied, especially in those areas that are used for economic activities, the proposed method provides tools for policy- and decision- makers to set science-based criteria for the delimitation of conservation areas.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.seares.2013.03.002>.

Acknowledgements

This study was supported by a contract undertaken between the Basque Water Agency–URA and AZTI-Tecnalia; likewise by the Ministry of Science and Innovation of the Spanish Government (Project Ref.: CTM2011-29473). We wish to thank: the Directorate-general for Public Works and Water Management, Division Zeeland of the Ministry of the Infrastructure and Environment (The Netherlands); the Radboud University (Nijmegen, The Netherlands); Bert Brinkman from the Institute for Marine Resources and Ecosystem Studies (The Netherlands); and Annette Wielemaker from the Royal Netherlands Institute for Sea Research (NIOZ). M. Valle has benefited from a PhD Scholarship granted by the Iñaki Goenaga – Technology Centres Foundation. We wish to thank also Professor Michael Collins (School of Ocean and Earth Science, University of Southampton (UK), Estación Marina de Plentzia (University of the Basque Country) and AZTI-Tecnalia (Spain)), for kindly advising us on some details of the manuscript. The comments of two anonymous reviewers have improved considerably the first manuscript draft. This paper is contribution number 617 from AZTI-Tecnalia (Marine Research Division).

References

- Baretta, J.W., Ruardij, P., 1988. Modelling the ecosystem of the Ems Estuary. In: Baretta, J.W., Ruardij, P. (Eds.), Tidal flat estuaries. Simulation and analysis of the Ems estuary. Ecological Studies, vol. 71. Springer, Berlin, pp. 3–13.
- Barillé, L., Robin, M., Harin, N., Bargain, A., Launeau, P., 2010. Increase in seagrass distribution at Bourgneuf Bay (France) detected by spatial remote sensing. *Aquatic Botany* 92, 185–194.
- Bekkby, T., Rinde, E., Erikstad, L., Bakkestuen, V., Longva, O., Christensen, O., Isæus, M., Isachsen, P.E., 2008. Spatial probability modelling of eelgrass (*Zostera marina*) distribution on the west coast of Norway. *ICES Journal of Marine Science* 65, 1093–1101.
- Bell, S.S., Robbins, B.D., Jensen, S.L., 1999. Gap dynamics in a seagrass landscape. *Ecosystems* 2, 493–504.
- Bisrat, S.A., White, M.A., Beard, K.H., Richard Cutler, D., 2012. Predicting the distribution potential of an invasive frog using remotely sensed data in Hawaii. *Diversity and Distributions* 18, 648–660.
- Björk, M., Short, F., Mcleod, E., Beer, S., 2008. Managing Seagrasses for Resilience to Climate Change. IUCN, Gland, Switzerland 56.
- Bos, A.R., van Katwijk, M.M., 2007. Planting density, hydrodynamic exposure and mussel beds affect survival of transplanted intertidal eelgrass. *Marine Ecology Progress Series* 336, 121–129.
- Bos, A.R., Dankers, N., Groeneweg, A.H., Carin, D., Hermus, R., Jager, Z., de Jong, D.J., Smit, T., de Vlas, J., van Wieringen, M., van Katwijk, M.M., 2005. Eelgrass (*Zostera marina* L.) in the western Wadden Sea: monitoring, habitat suitability model, transplantations and communication. In: Herrier, J.L., Mees, J., Salman, A., Seys, J., Van Nieuwenhuysse, H., Dobbelaere, I. (Eds.), Proceedings ‘Dunes and Estuaries 2005’—International Conference on Nature Restoration Practices in European Coastal Habitats: VLIZ Special Publication, 19, pp. 95–109. Koksijde, Belgium; xiv + 685.
- Breiman, L., 2010. Random forest. *Machine Learning* 45, 15–32.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37, 35–46.
- Costanza, R., d’Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O’Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M., 1997. The value of the world’s ecosystem services and natural capital. *Nature* 387, 253–260.
- Cunha, A.H., Marbá, N.N., van Katwijk, M.M., Pickerell, C., Henriques, M., Bernard, G., Ferreira, A., Garcia, S., Garmendia, J.M., Manent, P., 2012. Changing paradigms in seagrass restoration. *Restoration Ecology* 20, 427–430.
- Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792.
- de Jonge, V.N., 1992a. Tidal flow and residual flow in the Ems estuary. *Estuarine, Coastal and Shelf Science* 34, 1–22.
- de Jonge, V.N., 1992b. Physical processes and dynamics of microphytobenthos in the Ems estuary (The Netherlands). Ph.D. Thesis, University of Groningen, pp 176.
- de Jonge, V.N., 2000. Importance of temporal and spatial scales in applying biological and physical process knowledge in coastal management, an example for the Ems estuary. *Continental Shelf Research* 20, 1655–1686.
- den Hartog, C., 1971. The dynamic aspect in the ecology of sea-grass communities. *Thalassia Jugoslavica* 7, 101–1012.
- Dolch, T., Buschbaum, C., Reise, K., 2012. Persisting intertidal seagrass beds in the northern Wadden Sea since the 1930s. *Journal of Sea Research* 82, 134–141.
- Duarte, C.M., 2002. The future of seagrass meadows. *Environmental Conservation* 29, 192–206.
- Duarte, C.M., Fourqurean, J.W., Krause-Jensen, D., Olesen, B., 2006. Dynamics of seagrass stability and change. In: Larkum, A.W.D., Orth, R.J., Duarte, C.M. (Eds.), *Seagrasses: Biology, Ecology and Conservation*. Springer, The Netherlands, pp. 271–294.
- Elith, J., Leathwick, J.R., 2009. Conservation prioritization using species distribution models. In: Moilanen, A., Wilson, K.A., Possingham, H.P. (Eds.), *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools*. Oxford University Press, Oxford, pp. 70–93.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., 2006. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography* 29, 129–151.
- Erftemeijer, P.L.A., 2005. Trendanalyse van zeegrasverspreiding in de Nederlandse Waddenzee 1988–2003. WL Delft Hydraulics 105.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24, 38–49.
- Fonseca, M.S., Bell, S.S., 1998. Influence of physical setting on seagrass landscapes near Beaufort, North Carolina, USA. *Marine Ecology Progress Series* 171, 109–121.
- Fonseca, M.S., Kenworthy, W.J., Whitfield, P.E., 2000. Temporal dynamics of seagrass landscapes: a preliminary comparison of chronic and extreme disturbance events. *Biologia Marina Mediterranea* 7, 373–376.
- Franklin, J., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge 320.
- Gahegan, 2003. Is inductive machine learning just another wild goose (or might it lay the golden egg)? *International Journal of Geographical Information Science* 17, 69–92.
- Green, E.P., Short, F.T., 2003. *World Atlas of Seagrasses*. University of California Press, Berkeley 324.
- Greve, T.M., Binzer, T., 2004. Which factors regulate seagrass growth and distribution? In: Borum, J., Duarte, C.M., Krause-Jensen, D., Greve, T.M. (Eds.), *European seagrasses: An introduction to monitoring and management: The M&MS project*. ISBN: 87-89143-21-3 (Internet version available at <http://www.seagrasses.org>).
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, 993–1009.
- Guisan, A., Zimmermann, N., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147–186.
- Guisan, A., Edwards Jr., T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157, 89–100.
- Han, Q.Y., Bouma, T.J., Brun, F.G., Suykerbuyk, W., van Katwijk, M.M., 2012. Resilience of *Zostera noltii* to burial and erosion disturbances. *Marine Ecology Progress Series* 449, 133–143.

- Hastie, T., 2011. GAM: generalized additive models. R package version 1.06.2. <http://CRAN.R-project.org/package=gam>.
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2012. Dismo: species distribution modelling. R package version 0.7-17. <http://CRAN.R-project.org/package=dismo>.
- Hughes, R., Willians, S., Duarte, C.M., Heck Jr., K.L., Waycott, M., 2009. Associations of concern: declining seagrasses and threatened dependent species. *Frontiers in Ecology and the Environment* 7, 242–246.
- Jager, Z., Bartels, W., 2002. Optimale zoetwateraanvoer naar de Waddenzee. Werkdocument RIKZ/AB/2002.604x. Haren, The Netherlands.
- Kamerlings, P., Hemminga, M.A., de Jong, D.J., 1999. Significance of salinity and silicon levels for growth of a formerly estuarine eelgrass (*Zostera marina*) population (Lake Grevelingen, The Netherlands). *Marine Biology* 133, 527–539.
- Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab — an S4 package for kernel methods in R. *Journal of Statistical Software* 11, 1–20.
- Keddy, C.J., Patriquin, D.G., 1978. An annual form of eelgrass in Nova Scotia. *Aquatic Botany* 5, 163–170.
- Kirkman, H., 1997. Seagrasses of Australia. Australia: State of the Environment Technical Paper Series (Estuaries and the Sea). Department of the Environment, Canberra, Australia 36.
- Krause-Jensen, D., Almela, E.D., Cunha, A.H., Greve, T.M., 2004. Have seagrass distribution and abundance changed? In: Borum, J., Duarte, C.M., Krause-Jensen, D., Greve, T.M. (Eds.), *European seagrasses: an introduction to monitoring and management. The M&MS project*. ISBN: 87-89143-21-3 (Internet version available at <http://www.seagrasses.org>).
- Liaw, A., Wiener, M., 2002. Classification and regression by random forest. *R News* 2, 18–22.
- Liu, C., White, M., Newell, G., 2011. Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography* 34, 232–243.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17, 145–151.
- MacArthur, R.H., Wilson, E.O., 1967. *The Theory of Island Biogeography*. Princeton University press, Princeton, NJ 203.
- Martinet, J.F., 1782. Verhandeling over het wier der Zuiderzee. *Verhandelingen Hollandsche Maatschappij der Wetenschappen* 20, 54–129 (+ plates I–VI).
- Milborrow, S. Derived from mda: mars by Hastie, T., Tibshirani, R., 2012. earth: multivariate adaptive regression spline models. R package version 3.2–3. <http://CRAN.R-project.org/package=earth>.
- Narumalani, S., Jensen, J.R., Althausen, J.D., Burkhalter, S., Mackey, H.E., 1997. Aquatic macrophyte modelling using GIS and logistic multiple regression. *Photogrammetric Engineering and Remote Sensing* 63, 41–49.
- Oppel, S., Meirinho, A., Ramirez, I., Gardner, B., O'Connell, A., Louzao, M., 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation* 156, 94–104.
- Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133, 225–245.
- Phillips, J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling* 190, 231–259.
- Raes, N., ter Steege, H., 2007. A null-model for significance testing of presence-only species distribution models. *Ecography* 30, 727–736.
- Segurado, P., Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31, 1555–1568.
- Short, F.T., Koch, E.W., Creed, J.C., Magalhaes, K.M., Fernández, E., Gaeckle, J.L., 2006. SeagrassNet monitoring across the Americas: case studies of seagrass decline. *Marine Ecology* 27, 277–289.
- Talke, S.A., de Swart, H.E., 2006. Hydrodynamics and morphology in the Ems/dollard estuary: review of models, measurements, scientific literature, and the effects of changing conditions. Institute for Marine and Atmospheric Research Utrecht (IMAU) Report # R06-01.
- Trémolières, M., 2004. Plant response strategies to stress and disturbance: the case of aquatic plants. *Journal of Bioscience* 29, 461–470.
- Valle, M., Borja, Á., Chust, G., Galparsoro, I., Garmendia, J.M., 2011. Modelling suitable estuarine habitats for *Zostera noltii*, using Ecological Niche Factor Analysis and Bathymetric LiDAR. *Estuarine, Coastal and Shelf Science* 94, 144–154.
- van Katwijk, M.M., Hermus, D.C.R., 2000. The effect of physical disturbance on *Zostera marina* L. transplantations in the Dutch Wadden Sea. *Marine Ecology Progress Series* 208, 107–118.
- van Katwijk, M.M., Schmitz, G.H.W., Gasseling, A.P., van Avesaath, P.H., 1999. Effects of salinity and nutrient load and their interaction on *Zostera marina*. *Marine Ecology Progress Series* 190, 155–165.
- van Katwijk, M.M., Hermus, D.C.R., de Jong, D.J., de Jonge, V.N., 2000. Habitat suitability of the Wadden Sea for *Zostera marina* restoration. *Helgolander Marine Research* 54, 117–128.
- van Katwijk, M.M., Geerling, G.W., Rasin, R., van 't Veer, R., Bos, A.R., Hermus, D.C.R., van Wieringen, M., Jager, Z., Groeneweg, A.H., Erfteimeijer, P.L.A., van der Heide, T., de Jong, D.J., et al., 2006. Macrophytes in the western Wadden Sea: monitoring, invasion, transplantations, dynamics and European policy. In: Laursen, K., Marencic, H. (Eds.), *Proceedings 11th International Scientific Wadden Sea Symposium*, Esbjerg, pp. 89–98.
- van Katwijk, M.M., Bos, A.R., de Jonge, V.N., Hanssen, L.S.A.M., Hermus, D.C.R., de Jong, D.J., 2009. Guidelines for seagrass restoration: importance of habitat selection and donor population, spreading of risks, and ecosystem engineering effects. *Marine Pollution Bulletin* 58, 179–188.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, Fourth edition. Springer, New York 496.
- Waycott, M., Duarte, C.M., Carruthers, T.J.B., Orth, R.J., Dennison, W.C., Olyarnik, S., Calladine, A., Fourqurean, J.W., Heck, K.L., Hughes, A.R., Kendrick, G.A., Kenworthy, W.J., Short, F.T., Williams, S.L., 2009. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the National Academy of Sciences* 106, 12377–12381.
- Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M., Schwartz, M.W., 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* 15, 565–576.
- Zaniewski, A.E., Lehmann, A., Overton, J.M., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157, 261–280.